

DOCUMENT RESUME

ED 205 894

CS 006 170

AUTHOR Stetson, Elton G.
TITLE Reading Tests Don't Cheat, Do They?
PUB DATE [73]
NOTE 21p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; *Reading Achievement; Reading Comprehension;
*Reading Research; *Reading Tests; *Speed Reading;
Testing Problems; *Test Interpretation; *Test Theory;
Test Validity; Vocabulary

ABSTRACT

After employees of private firms completed several rapid reading classes and achieved remarkable gains on the Nelson-Denny Reading Test, the question was raised as to whether the increases in scores were due to the increased number of items attempted on the posttest. A preliminary analysis indicated that students attempted an average of 14.6 and 4.2 additional items on the vocabulary and comprehension tests respectively. Protocols of the posttest were rescored to determine percentile ranks on the same number of items that had been completed on the pretest. Percentile scores were then recorded for the pretest, posttest, and adjusted posttest for the 60 students in the study. Computations between various scores showed that the gains on the vocabulary test and on the comprehension test were due to the increase in the number of items attempted on the posttest. The results also indicated that when the posttest scores were adjusted to control for the number of items attempted, there were mean losses on the vocabulary and comprehension tests. The findings suggest that the validity of such tests to measure growth in a rapid reading course are highly suspect.
(Author/HOD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Elton G. Stetson

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Reading Tests Don't Cheat, Do They?

Elton G. Stetson

Introduction

Many of us who work in secondary and college level reading programs are involved in teaching reading improvement or study skills courses. One of the advantages of having such expertise is that local firms often call and request assistance in offering what they refer to as speed-reading classes. During the past year I have taught five such classes involving accountants, engineers, geologists, attorneys, and other well educated professionals.

On the first day of each class I picked up my box clearly marked SPEED READING, drove to the site, and delivered my famous introductory speech on the misconceptions associated with speed-reading instruction. During the second hour the Nelson-Denny Reading Test, Form C, was administered to everyone (Brown, 1973). The tests were then scored by the students, results posted in their folders, and the training sessions began. All classes met two hours each day for eight days.

On the final day of class, Form D of the same test was administered, scored, and compared with the results of the

ED205894

CS026170

pretest. In most cases students left the class with a lot of enthusiasm and ready to practice their new-found skills speed-reading their way to the top. I was happy too. The course evaluations were always good, and I knew that within 30 days there would be a check in the mail. The scores? Oh yes, the scores were excellent. You can speed read Table 1, and see for yourself that the test results were good. Not only did the 60 students increase their reading

Table 1 About Here

rate by 234 percent, they also increased their vocabulary and comprehension scores even though neither was emphasized. Individual analyses of variance computed between the pre- and post-test scores indicated significant gains at or below the .05 level of confidence on the reading rate and vocabulary tests. The gains in comprehension (+ 4.9 points) did not meet the criterion for significance (See F-values under Table 1). The goal of the course, to increase rate while maintaining comprehension, had been met.

Something Seemed Strange

It was during the third or fourth time through one of these classes when it dawned on me that a significantly

large number of students had completed the post-test who had not completed the pretest. I decided to investigate further by determining the number of students in all five classes who finished the pre- and post-test within the allotted time. In addition, the average number of items attempted on the pre- and post-test was calculated. The data in Table 2 illustrates a 30% increase in the number of students who completed the vocabulary post-test and a 42% increase in those who completed the comprehension

Table 2 About Here

post-test. There was also an increase in the items attempted (+14.6 for vocabulary; +4.2 for comprehension). Since time limits had been carefully followed, I immediately credited the increase to the effects of the training. After all, this was speed reading.

I should have stopped right there. However, the data in Table 2 increased my curiosity. Was it possible that the gains in vocabulary and comprehension were the result of more items attempted rather than an increase in vocabulary or comprehension ability?

Comparing Actual Vs. Adjusted Scores

To explore this question further I decided to go back to the post-test protocols of all 60 students in my five classes and calculate the percentile scores in two different ways. First, the percentile scores were calculated in the normal manner by counting the correct responses and converting raw scores to percentile ranks using the tables available in the manual. Second, the post-tests were rescored by counting the correct responses only as far in the test as the student had gone when the pretest was taken. In other words, if a student completed 50 items on the pretest and 65 items on the post-test, percentile scores were determined based on 65 items and on 50 items. This latter calculation will be referred to as "adjusted scores." For each student a pretest score, a post-test score, and an adjusted post-test score was calculated.

Actual Vs. Adjusted Vocabulary Growth

The mean percentile scores (pre-, post-, and adjusted post) for the five classes taking the vocabulary test are displayed in Table 3. A comparison of the pre- and

Table 3 About Here

post test scores (column a and b) shows that all classes

improved with an overall increase of 10.74 points. An analysis of variance (Downie & Heath, 1965) calculated between the scores indicated significance, $F(1,118) = 4.21$, $p < .05$. Therefore, 15 to 16 hours of instruction resulted in a significant increase in vocabulary.

Table 3A is optional but is not referred to in the script

A second analysis was completed between the mean scores on the pretest and adjusted post-test (column a and d). In this comparison there was a decrease of 1.7 percentile points on the adjusted post-test. The adjusted scores are those that would have been achieved on the post-test had the students completed the same number of items on the post-test that were completed on the pretest. An analysis of variance calculated between those two scores indicated non-significance, $F(1,118) = 1.08$, $p > .05$. Interpreted, the effects of the course resulted in a loss of achievement, though not significant, when the number of items attempted is held constant.

Table 3b is optional but is not referred to in the manuscript.

It appears that the gains in vocabulary scores could

be attributed to the increase in the number of items attempted. The losses achieved when the adjusted post-test scores were considered appear to be caused by a slight decrease in test efficiency. Test efficiency is determined by dividing the total number of correct items by the total number attempted. The mean efficiency for all students was 83.1% on the pretest, 82.4% on the adjusted post-test, and 79.3% on the actual post-test.

Of interest is the efficiency rating of 58.4% on the post-test items that were attempted beyond those attempted on the pretest. A total of 875 more items were attempted on the post-test, and only 511 of these were correct resulting in the low efficiency rating. Although each student averaged 14.6 additional items attempted on the post-test, only six or seven of these had to be correct in order to raise the overall percentile score by 10 points or more.

Perhaps those taking the post-test felt pressure during the final minutes of the test and began to respond more quickly, take more chances, and reduce their test efficiency.

Actual Vs. Adjusted Comprehension Growth

The identical procedures were followed with the comprehension scores. Table 4 displays the mean percentile scores (pretest, post-test, and adjusted post-test) for each class. There was actual growth in comprehension among

Table 4 About Here

four of the five classes with an overall gain of 4.9 percentile points (Columns a,b, and c). When the adjusted post-test scores were compared with the pretest scores, there was a loss of 4.5 points (columns a,d, and e). When separate analyses of variances were computed between the pre- and post-test means ($F = 1.14$), and between the pre- and adjusted post-test means ($F = .84$), none of the F-values were high enough to reach the alpha level needed for significance at the .05 level of confidence (3.94; $df = 1/118$ (1/100)).

Table 4a, 4b, and 4c are optional but are not referred to in the Manuscript

The increase in comprehension also appears to be attributed to the increase in the number of items attempted. Students attempted an average of 4.2 additional items on the post-test of which 1.5 were correct. However, the additional 1.5 correct answers counted as three points in the scoring system, and these three points accounted for the 9.4 difference between the scores on the actual post-test (45.6 %ile) and the adjusted post-test (36.2 %ile).

The test efficiency ratings on the comprehension test

dropped from 72.5 percent on the pretest to 66 percent on the post-test. The efficiency rating on the additional 4.2 items attempted on the post-test dropped drastically to 35.7 percent. As was the case with the vocabulary test, it seems that completing more items even at the expense of making more mistakes will produce higher scores.

Implications

There are several implications which are presented here to provoke further discussion.

1. It is suggested that the effect of the rapid reading course, by its very nature, contributed to the increase of 14.6 items attempted on the vocabulary post-test. This increase in items attempted is also the primary influence on the significant increase in vocabulary (+10.74 points; $p = < .05$). Had the students attempted the same number of items on both the pre- and post-tests, there would have been a decrease in scores as the adjusted post-test scores indicate. It is likely that the training influenced students to be more aware of speed which increased the number of items they attempted, lowered their test efficiency, but resulted in higher scores. This produced a false impression of growth when the same tests, administered without a time limitation, may have produced losses on the post-test.

2. An average of 4.2 additional items were attempted on the comprehension post-test. This seems somewhat confusing because of the spectacular increase of 234 percent in reading rate. This increased rate should have resulted in the completion of more than 4.2 additional items. Apparently this high reading rate, determined from the first passage, did not sustain itself throughout the comprehension test. Perhaps the awareness of timing on the first passage, coupled with a built-in desire to improve the low pretest scores, resulted in spuriously high rate scores on the post-test. Even if the same post-test rate of reading had been maintained during the reading of all eight passages, the students would have likely taken a great deal more time to ponder questions or reread portions of the passages, a sure indication that the high rate may have also produced a false impression of growth.

3. Directions for the comprehension test allow students to refer to the passages while answering the questions. At the pretest level when the majority of students did not finish the test, very little time could have been spent rechecking answers or referring back to the passage. During the post-test, 80 percent of the students finished ahead of time. Therefore, many more had opportunities to reread and change answers. According to their own feedback, most of those who finished the test ahead of time did return for a second

reading of some portion of the test. They also claimed they usually returned to the eight questions over the first passage because they felt more unsure of their answers to passage I than other passages. This is further support that the rate of reading on the post-test is unrealistically high. Had directions prohibited a rereading of the passages, there is little question that there would have been a sharp decrease in scores.

In essence, the results of this study suggest that the use of the Nelson-Denny Reading Test in a rapid reading course for pre-post-test analysis may be unrealistic and an invalid way of determining growth. The scores on the timed vocabulary and comprehension tests are easily increased simply by attempting more items on the post-test than on the pretest. In some cases, one additional correct response can account for increases of up to eight percentile points. Most students who take courses involving rapid reading techniques will naturally want to demonstrate their new abilities on the post-test. What is not known is that the higher scores may have been created simply by attempting more items even though their test efficiency may have been greatly reduced.

Furthermore, there is no check on the effectiveness of the reading rate score. A student can obtain a spuriously high reading rate, fail all eight questions on the passage

on which the rate is determined, and still achieve an increase in comprehension on the post-test. The combination of attempting more items and having the privilege of rereading passages and changing answers can easily produce a false picture of reading rate and comprehension. This lack of control on the comprehension test causes any test analysis to be highly suspect, both for rate and for comprehension.

While I am not willing to generalize these findings outside of the present study, there are a number of questions that this study raises, the answers to which could have a profound impact on the use of this and other tests in courses involving rapid reading training. There is little doubt that the students in these classes learned a great deal about adjusting reading rate according to the purpose for reading. They also developed techniques for reading considerably faster than they could have previously. However, it seems apparent that those who use instruments such as the Nelson-Denny in rapid reading courses can virtually guarantee success.

Conclusions

The validity of the Nelson-Denny Reading Test is not being challenged nor is its association with other instruments, its use as a predictor of academic success, or its diagnostic

value in the development of instructional programs. At question is its use as a measure of growth and change, particularly when used in classes where rapid reading is taught. Perhaps the authors of the Nelson-Denny Reading Test as well as others who create similar instruments could explore the validity and reliability factors associated with equivalent form tests having the following features:

1. Vocabulary tests that are untimed and designed to be finished by all students.
2. Comprehension tests that would not allow for the rereading of passages, particularly those over which a reading rate might be calculated.
3. Reading rate calculations that would involve more than one passage and more than one minute. These multiple readings could then be averaged for a more realistic rate score. For example, one instructor asked her students to "mark" at the 30, 60, 90 and 120 second intervals. Each reading was converted to a word-per-minute equivalent and then averaged. Reading rate measures without a check on comprehension are questionable.
4. A composite score computed from the interaction of the comprehension and reading rate scores, similar to the composite score currently

available for the combined vocabulary and comprehension scores. This might control for spuriously high rates and spuriously low comprehension.

While reading tests may not cheat, they may not be totally honest.

References

- Brown, J.I., Nelson, M.J., & Denny, E.C. Examiner's manual:
The Nelson-Denny reading test. Boston: Houghton
Mifflin Company, 1973.
- Downie, N.M., & Heath, R.W. Basic statistical methods
(2nd. ed.). New York: Harper & Row, Publishers, 1965.

Table 1

Mean Pre- and Post-test Comparisons
For 60 Students in Five Rapid Reading Classes

	Rdg. Rate Wds/Min. $\bar{X}$	Rdg. Rate Percentile $\bar{X}$	Vocab. Percentile $\bar{X}$	Comp. Percentile $\bar{X}$
Pretest Form C	203	24.2	60.4	40.7
Post-test Form D	477	79.4	71.1	45.6
Difference	+ 274 (1)	+ 55.2 (2)	+ 10.7 (3)	+ 4.9 (4)

(1) $\underline{F}=27.2$; (df= 1,118) = 6.90; $p < .01$

(2) $\underline{F}=43.6$; (df= 1,118) = 6.90; $p < .01$

(3) $\underline{F}=4.21$; (df= 1,118) = 3.94; $p < .05$

(4) $\underline{F}=1.14$; (df= 1,118) = 3.94; $p > .05$

Table 2

Pre- and Post-test Comparisons of Items

Attempted and Tests Completed, N=60

	Vocabulary (1)			Comprehension (2)		
	Pre-	Post-	Diff.	Pre-	Post-	Diff.
$\bar{X}$ Items Attempted	66.8	81.4	+14.6	30.7	34.9	+ 4.2
Number Finishing All Items	4 (7%)	22 (37%)	+18 (30%)	23 (38%)	48 (80%)	+ 25 (42%)

(1) Total Possible = 100

(2) Total Possible = 36

Table 3
Mean Percentile Scores -- Pretest, Post-test,
Adjusted Post-test -- Five Classes, N=60

Class	N	(a) Pre- test $\bar{X}$	(b) Post test $\bar{X}$	(c) Diff. (b-a)	(d) Adjust. Post $\bar{X}$	(e) Adjust. Diff. (d-a)
1	13	33.7	37.3	+ 3.6	28.2	- 5.5
2	11	64.8	79.1	+14.3	63.9	- .9
3	13	64.0	76.8	+12.8	63.4	- .6
4	12	72.5	86.3	+13.8	71.5	- 1.0
5	11	67.0	76.2	+ 9.2	66.5	- .5
Grand Mean	60	60.4	71.1	+10.7	58.7	- 1.7

Table 4
Mean Comprehension Percentile Scores -- Pretest,
Post-test, and Adjusted Post-test -- Five Classes, N=60

Class	N	(a) Pre- test $\bar{X}$	(b) Post- test $\bar{X}$	(c) Diff. +/- (b-a)	(d) Adjust. Post $\bar{X}$	(e) Adjust. Diff. (d-a)
1	13	16.2	29.1	+ 2.9	16.2	-10.0
2	11	40.1	42.6	+ 2.5	35.7	- 4.4
3	13	45.2	66.1	+20.9	51.3	+ 6.1
4	12	48.7	45.8	- 2.9	41.2	- 7.5
5	11	43.5	44.6	+ 1.1	36.5	- 7.0
Grand Mean	60	40.7	45.6	+ 4.9	36.2	- 4.5

Table 3a

Analysis of Variance Between Pre- and Post-test
Vocabulary Percentile Scores, N = 60

Source of Variation	df	Sum of Squares	Mean Square	<u>F</u> *	p
Between Groups	1	3,392	3,392.	4.21	< .05
Within Groups	118	94,986	805		
Total	119	98,378			

*F (df = 1,118 (1,100)) = 3.94 at .05

Table 3b

Analysis of Variance Between Pre- and Adjusted Post-test
Vocabulary Percentile Scores, N = 60

Source of Variation	df	Sum of Square	Mean Square	<u>F</u> *	p
Between Groups	1	96	96	.108	NS
Within Groups	118	104,807	888		
Total	119	104,903			

*F (df = 1,118 (1,100)) = 3.94 at .05

Table 3c

Analysis of Variance Between Post-test and Adjusted
Post-test Vocabulary Percentile Scores, N = 60

Source of Variance	df	Sum of Squares	Mean Square	<u>F</u> *	p
Between Groups	1	4,625	4,625	5.32	< .05
Within Groups	118	102,717	870		
Total	119	107,342			

*F (df = 1,118 (1,100)) = 3.94 at .05

Table 4a

Analysis of Variance Between Pre- and Post-test
Comprehension Percentile Scores, N = 60

Source of Variation	df	Sum of Squares	Mean Square	$\underline{F}^*$	p
Between Groups	1	775	775	1.14	NS
Within Groups	118	80,104	679		
Total	119	80,879			

* $\underline{F}$ (df = 1,118 (1,100)) = 3.94 at .05

Table 4b

Analysis of Variance Between Pretest and Adjusted
Post-test Comprehension Percentile Scores, N = 60

Source of Variation	df	Sum of Squares	Mean Square	$\underline{F}^*$	p
Between Groups	1	634	634	.84	NS
Within Groups	118	89,078	755		
Total	119	89,712			

* $\underline{F}$ (df = 1,118 (1,100)) = 3.94 at .05

Table 4c

Analysis of Variance Between Post-test and Adjusted
post-test Comprehension Percentile Scores, N = 60

Source of Variance	df	Sum of Squares	Mean Square	$\underline{F}^*$	p
Between Groups	1	2,099	2,099	3.11	NS
Within Groups	118	79,634	675		
Total	119	81,733			

* $\underline{F}$ (df = 1,118 (1,100)) = 3.94 at .05